International Core Journal of Engineering Volume 8 Issue 7, 2022
ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0114

Multi-mode Segmentation Algorithm for Brain Tumor Using Deep Learning

Xiaoli Liu[a], Xiaorong Cheng[b]

Computer science department, North China Electric Power University, Baoding 071003, China

[a]Xiaoli_Liu19@163.com, [b]Xiaor_cheng@163.com

Abstract

Brain tumors, as malignant tumors that occur at all ages, are characterized by high lethality and difficult to cure, so early and accurate diagnosis can provide more support for disease prevention and treatment. To this end, an improved TransBTS model is proposed. Firstly, to address the problem of brain tumors with different shapes and sizes, the central cropping of each modal voxel block increases the network depth to obtain finer feature information; secondly, the residual convolution module with residual connections is substituted for ordinary convolution to retain more target features; finally, a hybrid attention mechanism is added in the process of feature extraction to suppress irrelevant target features and enhance the feature extraction capability in critical regions; using The Dice similarity coefficients can reach 85.43%, 84.76% and 81.84% in whole tumor, tumor core and enhanced tumor, respectively, tested with BraTS2021 dataset, which has better accuracy and better solves the current problem of heavy target segmentation task.

Keywords

Multi Mode; MRI; Brain Tumor Segmentation; Attention Mechanism.

1. Introduction

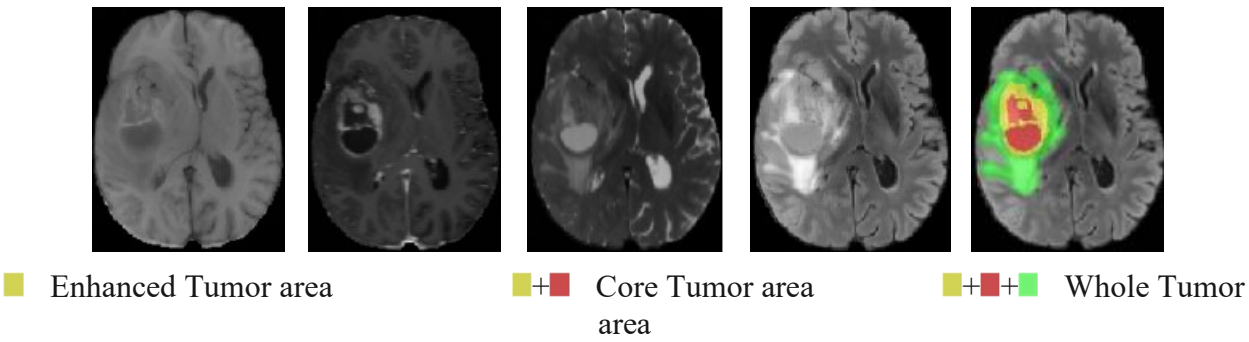


Fig. 1 Segmented slice maps of the four modalities of MRI

Brain tumor is a common malignant brain tumor, and glioma is among the most common primary brain tumors in adults, and its appearance varies in shape and size bringing different degrees of invasiveness to patients[1]. Therefore, accurate diagnosis of brain tumor is the crucial for surgical planning and subsequent protocol development. Currently, the main techniques to assist in brain tumor detection are CT and MRI. MRI is regarded clinically as the standard technique to assist in brain tumor scanning because it can compensate for the weakness of a single image modality that cannot adequately segment the tumor in the region of interest, has less radiation and has more advantages in terms of diagnostic information and resolution compared to CT images [2]. Medical

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0114

experts usually combine four modality images based on T1-weighted imaging (T1), T1-weighted with added contrast (T1c), T2-weighted imaging (T2) and fluid-attenuated inversion recovery image (FLAIR) scanned by MRI methods to determine the location areas of enhancing tumor, tumor core and whole tumor, as shown in Figure 1 for the four modalities in patient BraTS2021_00052 modal axial position slices and segmentation results in patient BraTS2021_00052.

However, relying solely on medical experts to manually label brain tumor lesions is not only inefficient, but also difficult to guarantee accuracy. Therefore, the use of computer-related algorithms for tumor assisted diagnosis can improve the tumor detection rate, assist physicians in diagnosis, and meet the market demand of surging patient data. The earliest segmentation methods were mainly based on machine learning algorithms, such as decision trees [3], random forests [4], but the complexity of the imaging preprocessing process and the drawback of being highly dependent on human experience made it difficult to be used for practical applications. Later, semi-automatic supervised learning algorithm models based on CNN neural networks emerged. Devorak [5] proposed a CNN segmentation method based on small image blocks to make the model more focused on the local structural features of images. Kamnitsas [6] used DeepMedic and CRF to capture 3D brain tumor features. Yet, none of them focused on global contextual features.The U-net network based on encoder-decoder structure proposed by Ronneberger [7] was first applied to Drosophila ventral nerve cord cell segmentation, followed by the 3D Unet model applicable to brain tumor segmentation proposed by HENRY [8], which can input full-volume images.Yu [9] used a full-convolution residual network with a depth of more than 50 layers convolutional residual network FCRN deep convolutional layers. Whereas, the pooling and downsampling of their proposed network during the encoding process can result in the loss of fine-grained information of the image. The self-attention mechanism of Transformer model can well compensate for the shortcomings of CNN by focusing attention on the global information. Therefore, Transformer combined with other convolutional networks became a high point after Unet. TransUnet [10] replaced a part of the Unet encoder with Transformer's attention format, using a slicing-based approach to slice 3D images into 2D pictures of the same size, which not only affects the segmentation accuracy but also relies on pre-trained model. TransBTS [11] is the first model that combines CNN and Transformer for 3D segmentation of brain tumors. 3D is sensitive to spatial information and can convolve the input 3D data in three directions (x, y, z) to extract features. Therefore, a neural network using 3D convolution can yield more results than a neural network using 2D convolution.

Based on the above analysis, this paper proposes a deep learning-based segmentation algorithm that uses the TransBTS model as a benchmark but extends and improves it in three ways: 1) to address the problem that the shape, size and location of brain tumors may vary from patient to patient, maximize the central cropping of the input voxel block size and increase the network depth to obtain more feature maps; 2) in the encoder and decoder sections using the residual convolution module to obtain better feature representation, and 3) incorporating a channel space mixing domain attention mechanism to suppress irrelevant features and focus on the key features useful for image segmentation.

2. Method

2.1 Network Architecture

Figure 2 shows the five-layer U-shaped network structure proposed in this paper. The network is improved on the basis of TransBTS structure and consists of encoder module, Transformer feature fusion module and decoder module.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0114

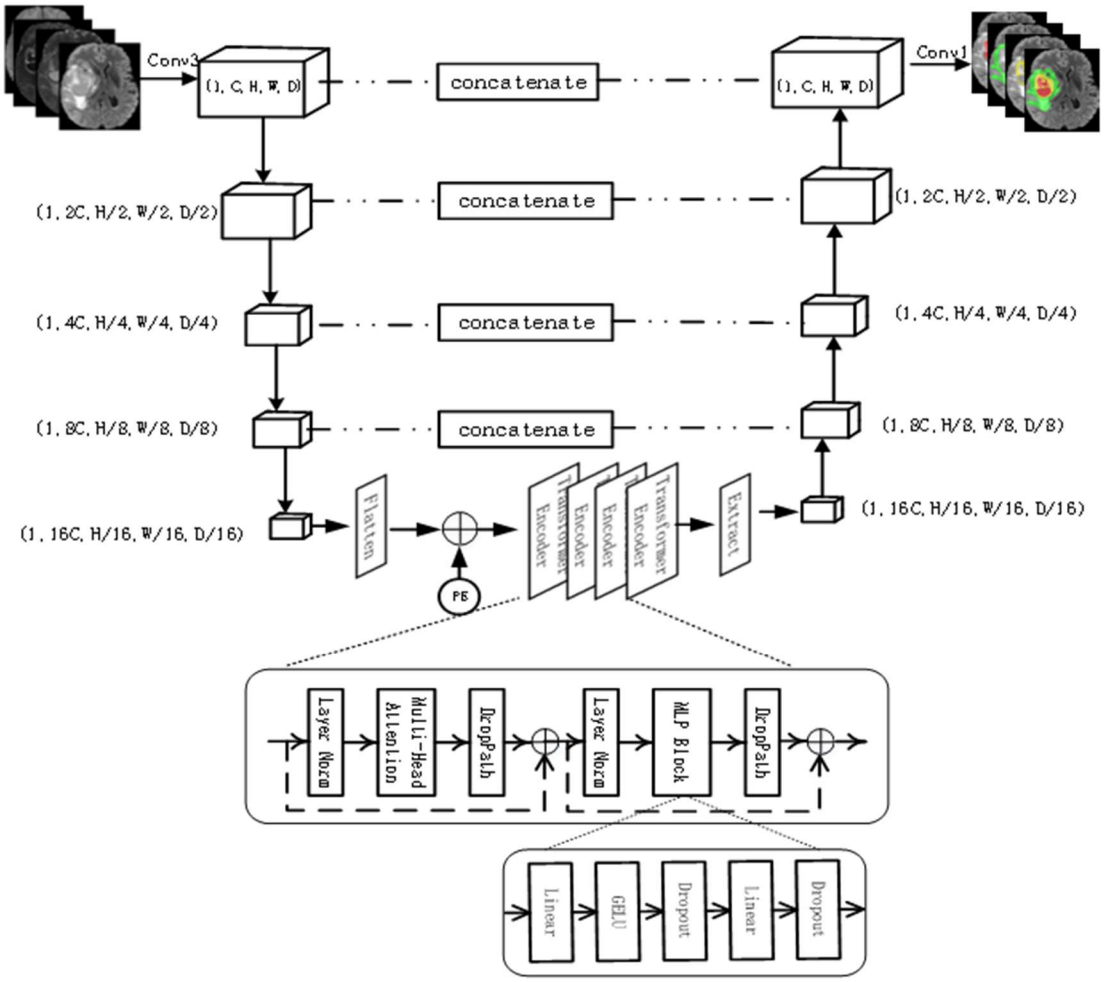


Fig. 2 Network architecture of improved TransBTS

The four modal brain tumor images are center cropped to obtain the feature maps (1,4,224,224,128) as the network model input, where the number of channels is 4 and the voxel block size is 224*224*128. After the convolution kernel size of 3*3*3 to obtain the feature map with 16 feature channels (1,16,224,224,128), the encoder process is formally started. The encoder module includes 5 residual connection network modules and 4 mixed domain attention mechanism downsampling operations. After encoding, the feature map is obtained as (1,16C,H/16,W/16,D/16), that is, (1,256,14,14,8). After the flatten operation, a one-dimensional sequence is captured and overlapped with the learnable position encoding as the input to the Vision Transformer encoder. The Vision Transformer encoder consists of four Transformer layers, and each layer of the Transformer layer consists of two modules, including the Multi-Head Attention (MHA) block and the MLP. Each modules are surrounded by LayerNorm and DropPath, and the modules are followed by residual connections to prevent important information. Among them, the MLP consists of two linear transformation functions Linear, GELU activation function and two Dropout. The decoding process uses jump joints to fuse low- and high-resolution feature information, and then layer by layer using a deconvolution operation to obtain the final high-resolution feature map (1,16,224,224,128). At last, a Conv1 convolution operation with a convolution kernel size of 1*1*1 and a softmax activation function is used to obtain the image size with four label channels as the output of the model. The specific flow of the model is shown in Table 1.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0114

Table 1. The specific process of the improved TransBTS network

Input: Feature Map(1,4,224,224,128)
(1) After the convolution kernel is 3×3×3 operation formed (1,C,H,W,D), that is (1,16,224,224,128) (2) Encoder a) residual convolution module b) downsampling c) hybrid attention mechanism (3) Vision Transformer network fuses global and local information (4) Decoder a) deconvolution upsampling b) Jump connection c) residual convolution module After the convolution kernel is 1*1*1 operation and sigmoid function to generate the feature weight map [0,1]
Output: Feature Map(1,4,224,224,128)

2.2 Residual Convolution Module

To ensure that the meaningful feature image content stays within the training region, we expand the input voxel block size and increase the network depth by one layer. However, the increase in network depth may introduce the risk of gradient degradation[12], so in this paper, the convolution operation of the encoder and decoder is replaced by a residual convolution module with residual connections. The residual convolution module is shown in Figure 3, with group normalization and PRelu instead of the regular BN and Relu.

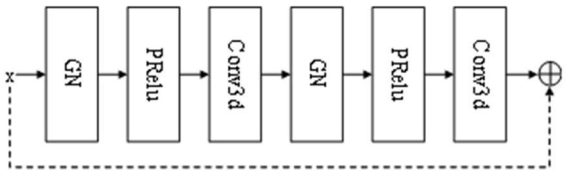


Fig. 3 Residual convolution module

On the one hand, group normalization divides the channels into groups, and then normalizes the channels using the mean and variance of each group[13]. It performs better than BatchNorm when the batch size is small (our batch size is 4). On the other hand, the PReLU with parameters[14], as shown in Eq. (1), $x_i > 0$, $P\mathrm{Re}lu(x_i) = x_i$, $x_i < 0$, $P\mathrm{Re}lu(x_i) = \alpha_i x_i$, and the slope of the negative part is based on the data rather than predefined. It is possible to solve the problem that the ReLU non-positive gradient is 0, resulting in some parameters never being updated. In this paper, we use α as 0.25.

$$P\mathrm{Re}lu(x_i) = \max(\alpha_i x_i, x_i) \tag{1}$$

2.3 Hybrid Domain Attention Mechanism

The attention mechanism in deep learning is similar to the human visual focus on a certain part of the region rather than the overall regional information, thus being better at grasping useful features. We use a hybrid attention mechanism in the channel domain and spatial domain for the downsampling process [15], as shown in Figure 4. The input feature map F, obtained by the channel attention module, is multiplied by F to obtain F'; F' is multiplied by the feature map F'' obtained by the spatial attention module and F' to obtain a new feature map for optimal adaptive feature map as the output.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0114

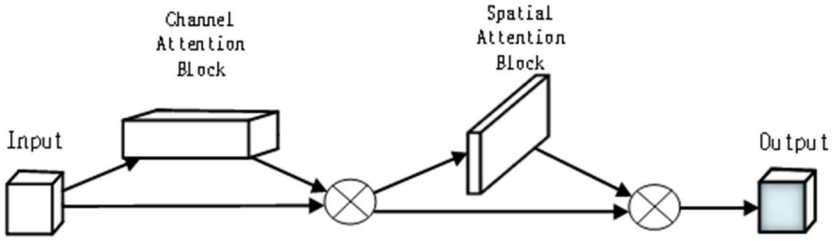


Fig. 4 CBAM attention mechanism

Specifically, a feature map F of size (C,H,W,D) is input, where C represents the number of channels, and H,W,D represent the height width and depth of voxel blocks, respectively. The (C,1,1,1) is obtained by averaging pooling and the (C,1,1,1) is obtained by maximum pooling, respectively, and then the number of channels is compressed by an MLP (Flatten-Linear-Relu-Linear). The obtained results are summed correspondingly and multiplied with F through a sigmoid function to obtain the output F' of the channel attention module. The output of the channel attention module is used as the input of the spatial attention module, and the mean and maximum values are first calculated along the channel dimension, and then the two are stacked together by the Concat operation to obtain a feature map with the number of channels of 2. Furthermore, the channel is changed to 1 by a convolution operation with a convolution kernel size of 7*7*7, a step size of 1, and a padding of 3, while keeping the image size unchanged. After a sigmoid function, it is multiplied with F' and transformed back to (C,H,W,D), and then connects with the initial input to the final output feature map. The experiment proves that the serial by channel attention mechanism and spatial attention mechanism can effectively solve the shortage of single attention mechanism for 3D brain tumor segmentation, and can achieve the effect of extracting more feature information of small area brain tumors and improving the segmentation results.

3. Experiment Research

3.1 Data Pre-processing

Medical Image Computing and Computer Assisted Intervention Society(MICCAL) has organized an annual multimodal brain tumor segmentation challenge since 2012 and published the corresponding MRI brain tumor segmentation public authority dataset. In this paper, the experimental data was selected from the BraTS2021 dataset [16][17][18], which contains 1251 training cases and 219 validation cases. The validation data contains only four modalities, while the training data contains four modal images and a gold standard image manually labeled by a neurologist. The gold standard image contains three tumor regions, namely: 1) Enhanced tumor region ET, which contains only enhanced tumor, labeled as 4, and generally appears only in HGG images. 2) Tumor core region, consisting of gangrenous NET labeled as 1 and enhanced tumor ET labeled as 4, which is the part to be removed during surgery. 3) Whole tumor region WT, consisting of gangrenous NET labeled as 1, the puffy region labeled as 2 and the enhanced tumor ET labeled as 4.

Considering the special characteristics of medical images, two pre-processing aspects of the dataset are done in this paper. On the one hand, 10% of the 1251 training data sets were randomly selected for test validation in the test set, using z-score intensity normalization to minimize the intensity variation within MRI of different patients and between multiple modalities of the same patient. On the other hand, the data were expanded using data enhancement methods including elastic deformation, mirror flip, Gaussian noise, and random intensity shifts.

3.2 Implementation Details

To evaluate the accuracy of brain tumor segmentation, Dice similarity coefficient (DSC) and Hausdorff distance (95%) [19] are used as evaluation criteria. The Dice coefficient measures the similarity between the predicted region and the real region, and the value ranges from [0,1], the larger

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0114

the value, the more similar the two sets are, and the Hausdorff (95%) distance mainly measures the matching degree of the boundary between the above two regions, the smaller the value, the more similar the two sets are.

The experimental platform is Windows 10 operating system, trained on CUDA10.2 architecture platform with 256GB memory, dual Intel(R) Xeon(R) Gold 5218 CPUs, Tesla V100*4 GPU server, network structure based on PyTorch framework, torch1.8 and python3.7.11 implementation. And the training is distributed with one machine and multiple cards.

3.3 Results

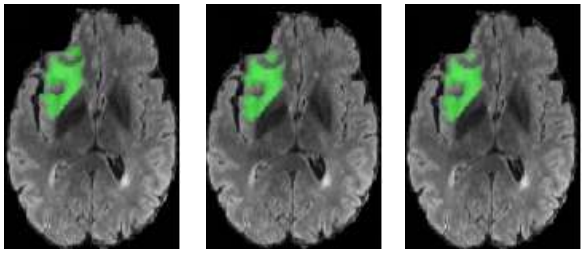
We trained the improved pre-model and the improved model for 300 rounds of each model, and each round took roughly about 25 min. After 300 training iterations, the model was tested and validated on 125 test sets of data, and the experimental results are shown in Table 2.

Table 2. Comparison of model segmentation results before and after improvement

Evaluation Metrics	Tumor category	Segmentation results		Comparison
		Pre-improvement model	Post-improvement model	
DSC	WT	0.850	0.854	+0.4%
	TC	0.844	0.847	+0.3%
	ET	0.817	0.818	+0.1%
HD95	WT	14.284	10.894	-3.39
	TC	11.290	9.577	-1.713
	ET	3.972	3.960	-0.012

As shown in the table, the Dice similarity coefficients of enhanced tumor, core tumor and whole tumor are improved by 0.1%, 0.3% and 0.4%, respectively, and the Hausdorff (95%) distances are shortened by 0.012, 1.713, and 3.39, respectively. The mean value of DSC for the three tumor segments improved by 0.27%, indicating that the tumor prediction results output by the model were more similar to the real results. The mean value of HD95 is shortened by 1.705, which further indicates that the predicted tumor regions are closer to the real regions.

Figure 5 shows the schematic diagram of the comparison of the actual results of model segmentation before and after the improvement for three randomly selected patient images, from the top to the bottom, the axial bitwise sections of the Flair modal images of patient BraTS2021_00466, patient BraTS2021_00518 and patient BraTS2021_01620, and from left to right, the real segmentation map, the TransBTS segmentation map and the improved TransBTS segmentation map. It can be seen that the proposed network in this paper can segment the whole tumor region, the tumor core and the enhanced tumor region more accurately. Compared with the original TransBTS, it is closer to the real segmentation map at the target edge.



ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0114

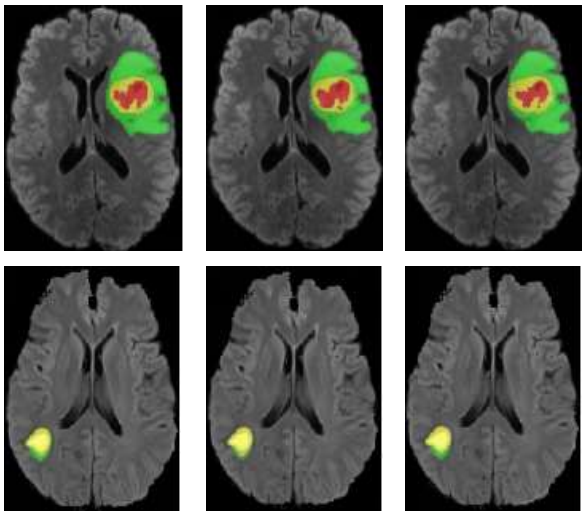


Fig. 5 Model segmentation effect comparison before and after improvement

Figure 6 shows the T2 modal images of the BraTS2021_00568 case, the first row is the original image of three slices in axial section, coronal section and sagittal section, and the second row is the resultant image segmented by the improved model, from left to right, the 78th slice of 155 slices in axial-view (78/155), the 121st slice of 240 slices in coronal-view (121/240) and the 121st slice of 240 slices in sagittal-view (121/240).

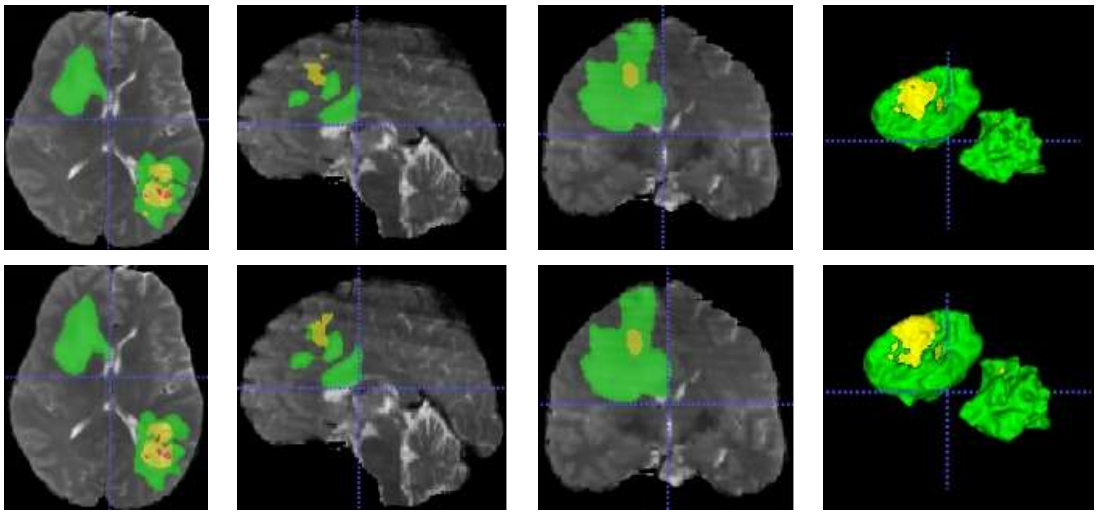


Fig. 6 ACS glioma segmentation result map of BraTS2021_00568

This patient suffers from two tumor blocks, as seen in three sections, and our improved model can accurately delineate the contours and boundaries of the enhancing tumor, the tumor core, and the whole tumor, which can assist the surgeon in further developing the surgical plan and subsequent treatment plan.

4. Conclusion

Segmenting brain tumors from multimodal images obtained from MRI has become a challenging task due to the specificity of the location of brain tumors and the diversity of tumor classifications. Compared with 2D series of network slicing segmentation brain tumor methods, 3D slicing methods based on deep learning can fuse more inter-modal information and obtain better performance. In the paper, a residual convolution operation with residual connections is added to address the problem that

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0114

the convolution operation of encoder and decoder may lose valid information, and a channel space mixing domain attention mechanism is added to the downsampling process of the encoder part. Tested and validated on the BraTS2021 dataset, the experimental results show that the improved network model can further improve the Dice similarity coefficient and reduce the HD95 distance for the three types of tumors. Therefore, the future will focus on the role and significance of various types of attention mechanisms in brain tumor segmentation tasks and medical segmentation tasks.

References

[1] K D Miller, M FidlerBenaoudia, K Theresa,et al. Cancer statistics for adolescents and young adults, 2020.[J]. CA: a cancer journal for clinicians,2020.

[2] R Wang, T Lei, R Cui, et al. Medical image segmentation using deep learning: A survey. IET Image Process. 16, 1243– 1267 (2022). https://doi.org/10.1049/ipr2.12419.

[3] D Zikic, B Glocker, E Konukoglu, et al. Decision Forests for Tissue-Specific Segmentation of High-Grade Gliomas in Multi-channel MR, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012. Springer,pp. 369–376.

[4] P Dollar, C L Zittnick. Structured forests for fast edge detection, International Conference on Computer Vision (ICCV), pp. 1841–1848 (2013).

[5] P Devorak, B Menze. Local structure prediction with convolutional neural networks for multimodal brain tumor segmentation, Proceedings of the International MICCAI Workshop on Medical Computer Vision, pp. 59–71, Cham, Germany, 2015.

[6] K Kamnitsas,C Ledig,V F Newcombe, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation[J]. Medical Image Analysis,2017,36.

[7] O Ronneberger, P Fischer, T Brox. U-net: Convolutional networks for biomedical image segmentation, arXiv: 1505.04597,2015.

[8] T Henry, A Carre, M Lerousseau, et al. Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks: a BraTS 2020 challenge solution[J].arXiv:2011.01045.2020.

[9] L Yu, H Chen, Q Dou, et al. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks, IEEE Transactions on Medical Imaging, vol. 36, no. 4, pp. 994-1004,2017.

[10] J N Chen, Y Y Lu, Q H Yu. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation.arXiv:2102.04306,2021.

[11] W Wang, C Chen , M Ding, et al. TransBTS: Multimodal Brain Tumor Segmentation Using Transformer[J]. arXiv:2103.04430,2021.

[12] K. He, X. Zhang, S. Ren, et al. Deep residual learning for image cognition,Proc. CVPR, pp. 770-778, 2016.

[13] Y Wu, K He,Group normalization, Proceedings of the European conference on computer vision (ECCV), p 3–19,2018.

[14] G D Hu, F Y Qian, L G Sha. Article Application of Deep Learning Technology in Glioma, J Healthc Eng, Volume 2022, Article ID 8507773.

[15] S Woo, J Park, J Lee, et al. CBAM: Convolutional block attention module, Proc. ECCV, pp. 3-19, 2018.

[16] U Baid, S Chodasara , S Mohan, et al. The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification[J]. arXiv: 2107.02314,2021.

[17] Menzeb, A Jakab, S Bauer, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)[J]. IEEE Transactions on Medical Imaging,2015,34(10), 1993-2024.

[18] S Bakas, H Akbari, A Sotiras, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features[J]. Nature Scientific Data,2017, 4:170117.

[19] M Soltaninejad, T Pridmore, M Pound. Efficient MRI brain tumor segmentation using multi-resolution encoder-decoder networks, Crimi A, Bakas S (eds) Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries. Springer, Cham, pp 30–39.